Non-Linear Identification on the Digital Magnetic Recording Channel

Nicholas P. Sands* and John M. Cioffi

Information Systems Laboratory
Stanford University
Stanford, CA 94305-4055

Abstract

We describe a method for identifying the digital magnetic recording channel in which data dependent nonlinear effects are present. The method can be extended to any discrete-input, discrete-time channel with finite memory. The channel will be considered as a finite state machine or lookup table, in which the number of states is related to the memory length of the channel. The model parameters can be derived using linear least-squares estimation or by adaptation of the model to minimize mean-squared error. A simple linear transformation (Hadamard transform) will be used to convert the finite state machine outputs to the coefficients of a Volterra Series model.

Prototype channels with various recording densities are studied with this technique.

The finite state machine description can be used to construct a trellis, from which receiver performance bounds can be derived.

1 Introduction

In discussions of signal detection schemes for digital magnetic recording, the channel has typically been considered as linear in the limited sense that superposition holds. As recording density is increased, this assumption breaks down, and nonlinear intersymbol interference is observed [1, 2]. The physical sources for this breakdown have been attributed to transition shifting during the write process and the prescence of previous magnetization in the medium [2, 3, 4].

Various models have been suggested to account for nonlinearities in this channel. The Volterra Series approach has been reported [5, 6] and the use of a finite state machine has also been discussed in [5] and [7].

In this paper, we present an identification procedure based on the finite state machine model. This model is general for a discrete-time channel with binary inputs and finite time response. The model parameters are derived using a least-squares technique, although adaptive methods could also be used. By appropriate arrangement of the model parameters, we show that a convenient linear transformation - the Hadamard transform[6] - yields the Volterra Series model for the channel. We then extract those Volterra

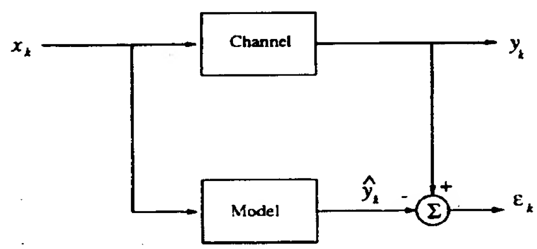

Figure 1: Channel Identification

Series kernels that have appreciable amplitude. The finite state machine(FSM) model is equivalent to a trellis description, and as such, the receiver performance bound for the channel under maximum likelihood sequence detection can be found.

2 Theory

2.1 Chip Decomposition

The channel output $y(t)$ is decomposed into signal chips [8], $y_k(t)$ each of duration T, defined by

$$y_k(t) = y(t + kT) \quad 0 \le t < T. \tag{1}$$

We can reconstruct $y(t)$ using

$$y(t) = \sum_k y_k(t - kT). \tag{2}$$

For practical purposes, we will deal with a sampled version of $y(t)$ at some arbitrary multiple ρ of the symbol rate $1/T$. The ith sample in the kth symbol period is then

$$y\,(t = (k + i/\rho)T) = y_{k,i}.$$

so that the chips are now vectors of length ρ, which are concatenated to form the sequence of channel output samples. We call this a *fractionally spaced* channel model. Note that we can let ρ grow to be arbitrarily large, in which case we have the continuous time function again.

*Supported by IBM Resident Study Program

a $1 \times 2^{\nu}$ row vector $\mathbf{x}_k^{(\nu)}$, a function of el input sequence $\{x_k\}$, defined recursively as follows

$$
\begin{aligned}
\mathbf{x}_k^{(0)} &= [1] \\
\mathbf{x}_k^{(1)} &= [1 \quad x_k] \\
\mathbf{x}_k^{(2)} &= [1 \quad x_k \quad x_{k-1} \quad x_k x_{k-1}] \\
&\vdots \\
\mathbf{x}_k^{(n+1)} &= \left[\mathbf{x}_k^{(n)} \mid x_{k-n}\mathbf{x}_k^{(n)}\right]
\end{aligned}
$$

for n up to $\nu - 1$. Also $\chi_{k,i}$ is the i-th element of $\mathbf{x}_k^{(\nu)}$. We will use the chip decomposition of (1) to express the model output as a sequence of chips $\hat{y}_k(t)$. We define the ν-order Volterra Series expansion for a channel of memory ν such that the kth chip is given by

$$\hat{y}_k(t) = \sum_{i=1}^{2^{\nu}} \chi_{k,i} v_i(t), \tag{3}$$

where $v_i(t)$ is a chip decomposition of the Volterra kernels of the channel.

Further, in the fractionally spaced channel model, the kth chip can be expressed as a vector $\hat{\mathbf{y}}_k$ of length ρ, given by

$$\hat{\mathbf{y}}_k = \mathbf{x}_k^{(\nu)}\mathbf{V}, \tag{4}$$

where the $2^{\nu} \times \rho$ matrix $\mathbf{V}$ consists of the chips of the 2^{ν} Volterra kernels, sampled at ρ/T.

2.3 Hadamard Transform

The Hadamard transformation can be used to map the FSM model parameters to the coefficients of the Volterra series model. There is a computationally efficient recursive method to do the transformation. The matrix is defined recursively

$$\mathbf{H}_0 = [1] \tag{5}$$

$$\mathbf{H}_{n+1} = \begin{bmatrix} \mathbf{H}_n & \mathbf{H}_n \\ \mathbf{H}_n & -\mathbf{H}_n \end{bmatrix} \tag{6}$$

The matrix has some useful properties: it is symmetric, as can be seen from the definition; and its inverse is just a scaled version of the matrix itself. Compute the product

$$\mathbf{H}_{n+1}\mathbf{H}_{n+1} = 2\begin{bmatrix} \mathbf{H}_n\mathbf{H}_n & 0 \\ 0 & \mathbf{H}_n\mathbf{H}_n \end{bmatrix} \tag{7}$$

repeating the recursion for n down to 0 for which

$$\mathbf{H}_0\mathbf{H}_0 = [1] \tag{8}$$

Thus,

$$\mathbf{H}_n\mathbf{H}_n = 2^n\mathbf{I} \tag{9}$$

and we can write the inverse of $\mathbf{H}$

$$\mathbf{H}_n^{-1} = \mathbf{H}_n/2^n \tag{10}$$

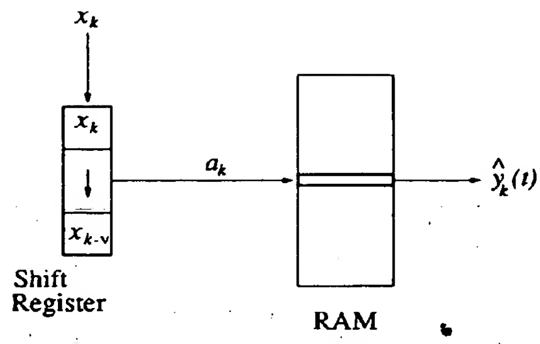

Figure 2: Finite state machine

2.4 Finite State Machine Model

In the case of saturation recording, the channel is constrained to have binary inputs, that is the field that is used to magnetize the medium takes one of two values. If we further assume that the memory of the channel is limited to at most ν symbols, we can ascribe the memory of past inputs to one of a finite number $2^{\nu-1}$ states. Then the channel has the Markov property, namely the next output depends on the current input and the current state. This can also be construed as a table in which the state and current input are combined into a binary address. Then this can be encoded into a integer that forms a pointer to a table entry. The channel outputs are the entries in the table.

The address pointer at the kth symbol will be denoted by $a_k^{(\nu)} \in \{0, \ldots, 2^{\nu} - 1\}$ and is computed using

$$a_k^{(\nu)} = \sum_{i=0}^{\nu-1} (1 + x_{k-i})\, 2^{i-1} \tag{11}$$

we can also establish the recursion

$$a_k^{(n+1)} = a_k^{(n)} + (1 + x_{k-n})\, 2^n \tag{12}$$

for $n = 0, 1, \ldots, \nu - 1$. We express the FSM channel model as a $2^{\nu} \times \rho$ matrix $\mathbf{R}$ with each row representing the ρ samples of the 2^{ν} FSM model chips.

Thus the kth output chip is given by

$$\hat{\mathbf{y}}_k = \mathbf{s}_k^{(\nu)}\mathbf{R} \tag{13}$$

where the vector $\mathbf{s}_k^{(\nu)}$ is a *selector* such that only one of its elements, namely the $a_k^{(\nu)}$th, is one and all other elements are zero.

Using the recursion in (12), we observe that

$$
\begin{aligned}
\mathbf{s}_k^{(n+1)} &= 2\left[\mathbf{s}_k^{(n)} \mid 0 \ldots 0\right] & x_{k-n} &= +1 \\
&= 2\left[0 \ldots 0 \mid \mathbf{s}_k^{(n)}\right] & x_{k-n} &= -1.
\end{aligned}
\tag{14}
$$

Given these relations, we will use the Hadamard matrix to compute

$$\mathbf{x}_k^{(n+1)}\mathbf{H}_{n+1} = \left[\mathbf{x}_k^{(n)}\mathbf{H}_n(1 + x_{k-n}) \;\mid\; \mathbf{x}_k^{(n)}\mathbf{H}_n(1 - x_{k-n})\right] \quad (15)$$

Thus

$$\begin{aligned}\mathbf{x}_k^{(n+1)}\mathbf{H}_{n+1} &= 2\left[\mathbf{x}_k^{(n)}\mathbf{H}_n \,|0\ldots 0\right] & x_{k-n} = +1\\ &= 2\left[0\ldots 0|\,\mathbf{x}_k^{(n)}\mathbf{H}_n\right] & x_{k-n} = -1\end{aligned}$$

We can see from these recursive relations that if $\mathbf{x}_k^{(n)}\mathbf{H}_n = \mathbf{s}_k^{(n)}$ then $\mathbf{x}_k^{(n+1)}\mathbf{H}_{n+1} = \mathbf{s}_k^{(n+1)}$. Then to complete the proof, set $n = 0$

$$\mathbf{x}_k^{(0)}\mathbf{H}_0 = [1][1] = [1] = \mathbf{s}_k^{(0)} \quad (16)$$

Thus we have an inductive proof that

$$\mathbf{x}_k^{(n)}\mathbf{H}_n = \mathbf{s}_k^{(n)} \quad (17)$$

for $n = 0, 1, \ldots, \nu$. Having established this relation we proceed to show the relation between the Volterra Series expansion and the FSM matrix $\mathbf{R}$.

Recall the Volterra Series expansion

$$\begin{aligned}\hat{\mathbf{y}}_k &= \mathbf{x}_k^{(\nu)}\mathbf{V} & (18)\\ &= \mathbf{x}_k^{(\nu)}\mathbf{H}\mathbf{H}^{-1}\mathbf{V}. & (19)\end{aligned}$$

Replacing $\mathbf{x}_k^{(\nu)}\mathbf{H}_\nu = \mathbf{s}_k^{(\nu)}$ and $\mathbf{H}^{-1} = \mathbf{H}/2^\nu$ we obtain

$$\hat{\mathbf{y}}_k = \mathbf{s}_k^{(\nu)}\mathbf{H}\mathbf{V}/2^\nu. \quad (20)$$

By equality with (13)

$$\begin{aligned}\mathbf{R} &= \mathbf{H}\mathbf{V}/2^\nu & (21)\\ \mathbf{V} &= \mathbf{H}\mathbf{R} & (22)\end{aligned}$$

Thus the Hadamard matrix is established as the relation between the Volterra kernels and the FSM model parameters.

2.5 Least-squares Identification

We consider the fractionally spaced samples of $y(t)$ as a set of ρ *subchannels* in the terminology of [9]. In this way, the row vector $\mathbf{y}_k$ consists of

$$\mathbf{y}_k = [y_{k,1} \;\; y_{k,2} \;\; \ldots \;\; y_{k,\rho}] \quad (23)$$

and the columns of matrix $\mathbf{R}$ are formed by subchannel models $\mathbf{r}_i$

$$\mathbf{R} = [\mathbf{r}_1 \;\; \mathbf{r}_2 \;\; \ldots \;\; \mathbf{r}_\rho] \quad (24)$$

We now assume a fixed value for ν in (13). Drop the superscript, and consider the estimate of the i-th such subchannel

$$\hat{y}_{k,i} = \mathbf{s}_k\mathbf{r}_i \quad (25)$$

We comp... channel outp...

$$e_{k,i} = y_{k,i} - \mathbf{s}_k\mathbf{r}_i$$

Although our model for the system is non-linear with respect to the original input sequence $\{x_k\}$, the error is linear with respect to the model parameters. Therefore we can use the results from linear estimation theory and find the set of model parameters $\mathbf{R}$ that minimizes the error sequence in the least squares sense.

Notice that we have in effect expanded our definition of *input*; from the sequence x_k to a nonlinear but completely defined function of the sequence $\mathbf{s}_k$ in accordance with (14), and in the process, retained linearity in the model parameters.

From linear estimation theory we invoke the Orthogonality Principle, which states that the error sequence for the i-th subchannel $e_{k,i}$ is orthogonal to the available data $\mathbf{s}_k$.

$$E\left[\mathbf{s}_k e_{k,i}\right] = 0 \quad (27)$$

In practice we can observe the channel's response to a finite input sequence, say of length N, so we estimate the expectation based on a finite amount of data. Thus we have a $N \times 2^\nu$ matrix $\mathbf{S}$ expressing the selector corresponding to each ν-tuple of the input sequence,

$$\mathbf{S} = \begin{bmatrix}\mathbf{s}_1\\ \mathbf{s}_2\\ \vdots\\ \mathbf{s}_N\end{bmatrix}, \quad (28)$$

and a $N \times 1$ column vector $\mathbf{y}_i$ containing the samples of the channel output. The estimate of the expectation then becomes a sum over the N sets of data, and can be expressed as the matrix product

$$\mathbf{S}'\mathbf{e}_i = 0 \quad (29)$$

Then since

$$\mathbf{e}_i = \mathbf{y}_i - \mathbf{S}\mathbf{r}_i \quad , \quad (30)$$

we obtain the minimum sample square error (MSSE) solution

$$\hat{\mathbf{r}}_{i,opt} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\mathbf{y}_i \quad (31)$$

and the MSSE is given by

$$\xi_{min,i} = (\mathbf{y}_i'\mathbf{y}_i - \mathbf{y}_i'\mathbf{S}\hat{\mathbf{r}}_{i,opt})/N \quad (32)$$

We can assemble the ρ subchannels together, and write

$$\mathbf{Y} = [\mathbf{y}_1 \;\; \mathbf{y}_2 \;\; \ldots \;\; \mathbf{y}_\rho] \quad (33)$$

$$\mathbf{E} = [\mathbf{e}_1 \;\; \mathbf{e}_2 \;\; \ldots \;\; \mathbf{e}_\rho]. \quad (34)$$

The complete FSM model is

$$\hat{\mathbf{R}}_{opt} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\mathbf{Y} \quad (35)$$

The minimum mean sample-squared error over all the ρ subchannels is given by $\sum_i \xi_{min,i}/\rho$, that is, the

, x .. E This expressed as

$$\xi_{min} = trace\left(\mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{S}\hat{\mathbf{R}}_{opt}\right)/\rho N \qquad (36)$$

Note that in practice to find $\mathbf{R}_{opt}$ it will not be necessary to compute the pseudo-inverse of $\mathbf{S}$, as implied by (35). Due to the sparse structure of $\mathbf{S}$, a special algorithm has been devised to compute $\hat{\mathbf{R}}_{opt}$ without matrix inversion.

2.6 Adaptive Identification

The use of an adaptive procedure for channel identification is suggested in [10]. Even though our model is nonlinear in nature, the error is linear with respect to the model parameters. The squared error will be quadratic, suggesting the use of a gradient search method, such as the LMS algorithm [10]. For the case of a linear model (FIR), the model parameters $\mathbf{w}_k$ are updated according to

$$\mathbf{w}_{k+1} = \mathbf{w}_k + 2\mu e_k \mathbf{x}_k \qquad (37)$$

Notice that at each iteration, all the parameters get updated.

For the FSM model, we have to modify the algorithm slightly [11] since the error is a nonlinear function of the channel input sequence x_k

$$r_{k,i} = r_{k,i} + 2\mu e_{k,i} \qquad (38)$$

In this case, only one of the model parameters receives an update. This leads to a slower convergence process, by a factor of $2^\nu/\nu$. This algorithm approaches the MMSE solution of (35), but with some excess error due to the gradient estimate.

3 Identification Procedure

3.1 Data Acquisition

The channel input sequence $\{x_k\}$ was obtained from a pseudo-random bit sequence (PRBS) rate $1/T$. The NRZI precode was applied - equivalent to $\frac{1}{1 \oplus D}$, where $\oplus$ denotes modulo-2 addition. This is a standard technique in digital magnetic recording and is used to cancel an inherent $1 - D$ factor in the readback process. It also has the desirable effect of making the output insensitive to signal inversions in the channel.

Next, the precoded binary sequence is modulated. The transmitter basis functions $\varphi(t)$ are defined as

$$\begin{aligned} \varphi(t) &= 1 \quad 0 \le t < T \\ &= 0 \quad otherwise \end{aligned}$$

The channel input is then

$$x(t) = \sum_k x_k \varphi(t - kT) \qquad (39)$$

where the sequence $\{x_k\} \in \{+1, -1\}$ is the pseudo-random sequence of period N.

normalized density	FIR $\nu = 8$	FIR $\nu = 256$	FSM $\nu = 8$
1.1	14.24	14.76	18.39
1.3	14.40	14.98	22.48
1.5	13.85	14.17	20.56
1.7	13.29	14.32	21.39
1.9	12.87	15.29	19.88

Table 1: SDR Results Summary for FIR and FSM models

The magnetic recording system used in this work consisted of a thin film disk and magnetoresistive(MR) head.

The PRBS was written at transition densities exceeding the nominal inner-diameter design density for this head and disk by factors up to 1.9.

Linear recording density is often characterized by the width of the step response at half its maximum amplitude, scaled to the symbol period - pw_{50}/T. The step responses on these channels were estimated and the pw_{50}/T ranged from 1.73 to 2.58.

A digital sampling oscilloscope was used to record the channel output. This instrument can digitize at a rate of 200×10^6 samples per second at 8 bits resolution and has a waveform memory of 40000 samples. This gives an oversampling ratio relative to the symbol rate of approximately 4, and with a sequence of length 1023, allows the response to nearly 10 repetitions of the input sequence to be digitized at once.

3.2 Post Processing

The digitized data sets undergo a number of processes to condition them for use in channel identification. A low-pass pre-filter was applied to filter some spurious high frequency noise. Resampling is necessary to re-create the frequency and phase of the write clock, since the digitizer was clocked asynchronously. Then, a cross-correlation technique was used to synchronize the read data with the written sequence. Noise averaging was done by averaging each repetition of the response to the sequence.

4 Results

The performance measure for modeling accuracy is residual mean-squared-error $E[(y_k - \hat{y}_k)^2]$. This is converted to ratio of signal to distortion (SDR), which is defined as

$$SDR = 10\log_{10}\left[\frac{E[y_k^2]}{E[(y_k - \hat{y}_k)^2]}\right] \qquad (40)$$

This is approximated from the results of (35) and (36) and using $E[y_k^2] = trace(\mathbf{Y}'\mathbf{Y})/\rho N$ we have

$$SDR = -10\log_{10}\left[1 - trace(\mathbf{Y}'\mathbf{S}\hat{\mathbf{R}}_{opt})/trace(\mathbf{Y}'\mathbf{Y})\right] \qquad (41)$$

The results of the channel identification process are shown in Table 1. Note that the FSM models give a clear 4-5dB gain over the linear models.

Having obtained the FSM model $\mathbf{R}_{opt}$, we now transform to a Volterra Series representation $\mathbf{V}$, via

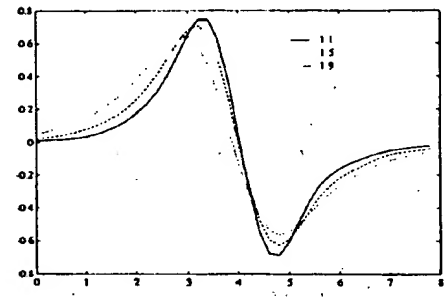

Figure 3: First order (linear) kernels.

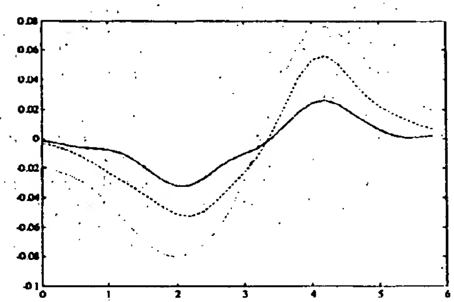

Figure 4: Kernels corresponding to $x_k x_{k-1} x_{k-2}$

(22). Since in most practical channels the types and extent of nonlinearity will be limited, many of the rows of V will be negligible, and we can focus on some specific types of nonlinearity. The first order kernel - the linear pulse response - is obtained by concatenating [8] those rows of V with row indices which are integral powers of 2. Other kernels are obtained in a similar way. Figure 3 shows the linear kernels, demonstrating increased dispersion as the recording density increases. The largest nonlinear kernel was the third-order nearest-neighbor corresponding to $x_k x_{k-1} x_{k-2}$, as shown in Figure 4. We interpret this interaction as evidence of the proximity-induced transition shift [2, 4].

5 Conclusion

We have observed that a purely linear model has limited accuracy on the magnetic recording channel at high density. Only 15dB of SDR was possible, even with a very long FIR model. The finite state machine model is equivalent to a complete Volterra Series expansion for these channels, given the discrete time, binary input sequence. In this sense the FSM model is general, if the channel is not time-varying. Using a 128 state model, we obtained SDR of about

model Another adv ... that it can be used directly to assess the receiv formance bound under maximum likelihood sequen detection (MLSD).

The disadvantages of such a model are involved with the number of parameters to be estimated(2^ν) and the fact that inspection of the parameter vector does not give any special insight into physical effects in the channel. We have therefore used the Hadamard matrix to transform the FSM parameter vector to the equivalent VS kernels. In this way, the higher order kernels can be inspected and the interactions from which they originate can be deduced.

References

[1] Paul Newby and Roger Wood. "The Effects of Nonlinear Distortion on Class IV Partial Response". *IEEE Transactions on Magnetics*, MAG-22:1203–1205, September 1986.

[2] D. Palmer, P. Ziperovich, R. Wood, and T. Howell. "Identification of Non-linear write effects using pseudorandom sequence". *IEEE Transactions on Magnetics*, MAG-23:2377–2379, September 1987.

[3] C. Michael Melas, Patrick Arnett, Irene Beardsley, and Dean Palmer. "Nonlinear Superposition in Saturation Recording of Disk Media". *IEEE Transactions on Magnetics*, MAG-23:2019–2021, September 1987.

[4] C. Tsang and Y-S. Tang. "Time-Domain Study of Proximity-Effect Induced Transition Shifts". *IEEE Transactions on Magnetics*, MAG-27:795–802, March 1991.

[5] O. Agazzi and D. G. Messerschmitt. "Nonlinear Echo Cancellation of Data Signals". *IEEE Transactions on Communications*, COM-30:2421–2433, November 1982.

[6] R. Hermann. "Volterra Modeling of Digital Magnetic Saturation Recording Channels". *IEEE Transactions on Magnetics*, MAG-26:2125–2127, September 1990.

[7] K.D. Fisher, J.M.Cioffi, and H.Thapar. "Modeling Thin-Film Storage Channels". *IEEE Transactions on Magnetics*, MAG-25(5), September 1989.

[8] Roger Wood. "Maximum Likelihood Reception on Nonlinear Dispersive Data Channels". *Electronics Letters*, 14:669–671, September 1978.

[9] J. M. Cioffi. "Least-squares Storage Channel Identification". *IBM Journal of Research and Development*, 30:310–320, May 1986.

[10] Bernard Widrow and Samuel D. Stearns. "*Adaptive Signal Processing*". Prentice-Hall, Englewood Cliffs, NJ, 1985.

[11] K. D. Fisher, J. M. Cioffi, W. L. Abbott, P. S. Bednarz, and C. M. Melas. "An Adaptive RAM-DFE for Storage Channels". To be published, 1991.